

Wireless Backhaul Networks: Capacity Bound, Scalability Analysis and Design Guidelines

Harpreet S. Dhillon and Giuseppe Caire

Abstract—This paper studies the scalability of a wireless backhaul network modeled as a *random extended network* with multi-antenna base stations (BSs), where the number of antennas per BS is allowed to scale as a function of the network size. The antenna scaling is justified by the current trend towards the use of higher carrier frequencies, which allows to pack large number of antennas in small form factors. The main goal is to study the per-BS antenna requirement that ensures scalability of this network, i.e., its ability to deliver non-vanishing rate to each source-destination pair. We first derive an information theoretic upper bound on the capacity of this network under a general propagation model, which provides a lower bound on the per-BS antenna requirement. Then, we characterize the scalability requirements for two competing strategies of interest: (i) *long hop*: each source-destination pair minimizes the number of hops by sacrificing multiplexing gain while achieving full beamforming (power) gain over each hop, and (ii) *short hop*: each source-destination pair communicates through a series of short hops, each achieving full multiplexing gain. While long hop may seem more intuitive in the context of massive multiple-input multiple-output (MIMO) transmission, we show that the short hop strategy is significantly more efficient in terms of per-BS antenna requirement for throughput scalability. As a part of the proof, we construct a scalable short hop strategy and show that it does not violate any fundamental limits on the spatial degrees of freedom (DoFs).

Index Terms—Wireless backhaul network, random extended network, capacity scaling, line-of-sight MIMO.

I. INTRODUCTION

To handle increasing wireless data traffic, cellular networks are undergoing a paradigm shift from a well-planned deployment of large tower-mounted and high power base stations (BSs) to an organic capacity-driven deployment of smaller and lower power BSs, often called *small cells* [3]. While conventional BSs are typically connected through a high capacity wired backhaul network, the same is not true for small cells, which may have to be densely deployed at more adverse locations [4]. For such systems, the backhaul represents one of the major bottlenecks in terms of throughput and deployment cost. A more economically viable alternative is to have wireless backhaul. The general idea is to establish high capacity wireless links that can carry cellular data from a cell to another cell and/or to some gateway node connected to the backbone network, possibly through multiple hops [5]. In this work, we focus on the design of throughput scalable

wireless backhaul systems, disregarding the cellular access links (uplink/downlink in each cell), which are assumed to operate in a different frequency band. As such, each BS is at the same time a source of traffic (generated by the users in its cell), a destination of traffic (to the users in its cell), and a relay, in the case of multi-hop backhaul systems.

The scaling of the per-connection (source-destination pair) throughput of wireless networks has been widely studied in the context of ad-hoc networks, which are conceptually closely related to the wireless backhaul system considered in this paper. While scalability is not achievable with single antenna transmission and multi-hop relaying [6], [7], there is definitely some hope in going to MIMO transmission. Besides, the current trend of moving towards higher and higher carrier frequencies (mm-wave communications) [8]–[10] makes it possible to pack more antennas in manageable form factors. For such systems, it is reasonable to envision a scenario where the number of antennas per BS, and hence the capacity of each backhaul link, increases with the network size. In this paper, we explore this networking paradigm in detail and show that it is possible to implement a scalable wireless backhaul network using MIMO transmission for each backhaul link, without violating any fundamental limits on the spatial DoFs.

As a disclaimer, we would like to point out that we are not envisaging a system where one should increase the number of antennas per BS with the addition of each new BS in the network in order to preserve scalability. Such a system would be highly impractical since it requires to change the hardware at each BS as the network topology evolves. In contrast, as often happens in Information Theory, our analysis captures an ensemble of systems designed such that the number of antennas per BS is a function of the overall network size. By capturing this dependency, we can provide useful design guidelines and answers to several important questions, such as, how many antennas per BS are necessary for a network of given size in order to achieve a certain desired target backhaul rate per source-destination pair.

A. Related Work

The study of the asymptotic throughput scaling laws of large wireless networks, where each node may be at the same time a source, a destination, and a relay node, has received a lot of attention in the past decade (e.g., see [6], [7], [11]–[15]). While giving a precise account of all these results is out of the scope of this work, we summarize briefly the relevant ones. For concreteness, we use the following *ordering* notation in this section: given two functions f and

This paper will be presented in part at IEEE ISIT, Honolulu, 2014 [1], [2]. The authors are with the Communication Sciences Institute (CSI), Department of Electrical Engineering, University of Southern California, Los Angeles, CA (email: {hdhillon; caire}@usc.edu).

This work was supported by NSF CCF-1161801 and a gift from Intel for research on 5G mm-wave wireless networks. Revised: June 12, 2014.

$g, f(n) = O(g(n))$ if there exists a constant c and integer N such that $f(n) \leq cg(n)$ for all $n > N$. Also, $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$. We will revisit this notation in Section II.

Studies on capacity scaling laws of wireless networks have considered both *extended* and *dense* network models, where the former refers to the deployment of nodes with constant spatial density, such that the network coverage area is $\Theta(n)$, with n denoting the number of nodes, while the latter refers to the case where the coverage area is $O(1)$ (constant with n), and the network density grows as $\Theta(n)$. In this paper we focus on a system where BSs are deployed with given (constant) density. Therefore, as far as the wireless backhaul is concerned, we are in the extended network regime. The seminal paper by Gupta and Kumar [6], successively refined by Franceschetti *et al.* [7], showed that the sum capacity of such networks scales as $\Theta(\sqrt{n})$ for single-antenna wireless networks under a multi-hop decode and forward relaying strategy. Since the number of source-destination pairs is $\Theta(n)$, this leads to a per-connection throughput that decreases as $\Theta(1/\sqrt{n})$. Successively (see [11] and the summary of results in [12]), it was shown that this behavior is information-theoretically order-optimal (over all possible relaying strategies) for power-law pathloss model with exponent $\alpha > 3$. For pathloss exponents $2 \leq \alpha < 3$, Özgür *et al.* [12] proposed a layered *hierarchical cooperation* scheme that can achieve sum throughput $\Theta(n^{1-\epsilon})$ for arbitrarily small $\epsilon > 0$ and a sufficiently large number of hierarchical layers. In contrast, by combining Maxwell propagation laws with an information theoretic cut-set bound argument, Franceschetti *et al.* [13] showed that the total spatial DoFs of the network are limited by $\Theta(\sqrt{a}/\lambda)$, where \sqrt{a} here can be interpreted as network “diameter”. Thus, for the extended model of area $a = \Theta(n)$, throughput $\Theta(\sqrt{n})$ is the best we can hope for. This *dichotomy* of results was settled in the case of line-of-sight (LoS) propagation by Lee *et al.* [14] and by Özgür *et al.* [15], which showed that these results hold in different operational regimes. The key insight is provided by the achievable DoFs (equivalently, spatial multiplexing gain) of point-to-point MIMO channels with LoS propagation, characterized in [14], [15]. In the case of non-LoS *dense scattering* propagation, Özgür *et al.* [16] have characterized different scaling regimes depending on the pathloss exponent and the typical nearest-neighbor SNR scaling exponent with n . These regimes hold as long as $n < \sqrt{a}/\lambda$, i.e., when the network operates far from the electromagnetic propagation bottleneck of [13].

In terms of the wireless backhaul network design, the focus of prior work has traditionally been on the cross layer optimization involving routing, scheduling and physical layer resource allocation. The main goal is to deliver data from the source BSs to a gateway node, which is connected to a wired backbone. A small sample of works in this direction is [17]–[19]. A more recent research direction has considered the performance of wireless backhaul links under practical considerations in the context of mm-wave communications. For instance, Hur *et al.* [20] studied mm-wave beamforming for wireless backhaul links with emphasis on the effect of wind induced pole movement on beam alignment. On the contrary, we consider a purely wireless backhaul network in this paper,

where none of the BSs have access to wired backhaul, and study its scalability as the number of BSs grow large. The main contributions are summarized next.

B. Novelty and Main Outcomes

Realistic setup for wireless backhaul networks. There are at least three main characteristics of modern wireless backhaul networks that are captured in this work: (i) the size of urban small cell networks is quickly growing, (ii) there is a gradual shift towards higher transmission frequencies, and (iii) increasing maturity of mm-wave communications makes it possible to support large number of antennas at each BS, thus forming high capacity backhaul links amongst BSs. We model this networking paradigm as an *extended* network, where both the transmission frequency and the number of antennas per BS can scale with the network size.

Information theoretic bound. In Section III, we derive an information theoretic upper bound on the capacity of the wireless backhaul network. The idea is to use information cut-set bound over a cut that divides the network into two equal halves. The information flow across the cut is upper bounded by the capacity of the MIMO channel, where the BSs on the either side of the cut act as a distributed transmitter and receiver, respectively. The main technical arguments are based on a simple generalization of *geometric exponential stripping* technique of [21] to channel matrices with complex-valued channel gains. An important consequence of this result is a lower bound on the number of antennas per BS required to ensure scalability.

Short hop vs. long hop strategies. In Sections IV and V, respectively, we consider two competing transmission strategies: (i) *long hop*: each source-destination pair minimizes the number of hops by sacrificing multiplexing gain and ideally achieving full power gain over each hop, and (ii) *short hop*: each source-destination pair communicates through a series of short hops, each achieving full multiplexing gain. While the long hop strategy may seem more reasonable, especially in the context of massive MIMO, where it is, in principle, possible to form thin beams in the direction of a far-off BS without creating excessive interference to its nearby BSs, we show that the short hop strategy is significantly efficient in terms of antenna requirement for throughput scalability. Hence, the relevance of works concerned with the *pointing problem* of narrow beams over long distances (long hops) in non-ideal conditions is, at best, questionable.

System design insights. The main design insight provided by our analysis is that it is possible to implement throughput-scalable wireless backhaul networks by forming high capacity MIMO links between BSs, without violating any fundamental limits on the electromagnetic propagation bottleneck. Under our achievable strategy, the capacity of each short hop MIMO link should have DoFs that scale as the square-root of the number of BSs. Quantitative concrete examples showing the attractiveness of the advocated approach are provided in Remark 3 and Section VI. In Section VI, we also demonstrate how our analysis can be extended to more general networks with some of the BSs having access to wired backhaul.

II. SYSTEM MODEL

We consider a cellular network where the locations of the BSs are sampled from a homogeneous Poisson Point Process (PPP) $\Phi \subset \mathbb{R}^2$ with density λ_b BSs per unit area. We further assume that none of the BSs have access to the conventional wired backhaul, and that all the data has to be communicated over wireless backhaul links. All the BSs share the same spectrum for their backhaul communication. As noted in the previous section, we do not consider the underlying cellular communication (uplink and downlink) between wireless users and BSs. Assuming a well-planned and load balanced system, we can easily imagine that each BS has to handle a fixed amount of traffic that does not grow with the overall number of BSs in the network. Our focus, here, is on the ability of a wireless backhaul network to relay such traffic.

For the scaling results, we consider the *random extended network model*, where we focus our attention on the box B_n with size $\sqrt{n} \times \sqrt{n}$. The number of BSs lying in B_n is a Poisson distributed random variable with mean $\lambda_b n$. We are concerned with the asymptotic capacity scaling of the network formed by the BSs inside B_n as $n \rightarrow \infty$. In this regime, the box B_n also grows, eventually encompassing all the points of Φ , hence the name “extended network”. Assuming uniform traffic, the source-destination pairs in B_n are picked uniformly at random, such that each BS is a destination of exactly one source.

For the wireless backhaul links, each BS has $\Psi(n)$ antennas, where $\Psi(\cdot) : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ is a monotonically non-decreasing function of n . Note that the antenna scaling assumption already appears in the literature, although in slightly different contexts, e.g., see [22]. Furthermore, we assume that the physical dimensions (size) of antenna array does not change with n . Denoting the distance between the k^{th} antenna of the transmitting BS to the i^{th} antenna of the receiving BS by d_{ik} , the baseband channel gain h_{ik} between these two antennas is

$$h_{ik} = \sqrt{l(d_{ik})} \exp(j\theta_{ik}), \quad (1)$$

where $l(d_{ik}) = \min\{1, d_{ik}^{-\alpha}\}$ is a bounded power-law pathloss function with exponent $\alpha > 2$, and θ_{ik} denotes phase rotation, which is typically a function of d_{ik} . Note that our analysis in Sections III and IV holds for any given $\{\theta_{ik}\}$, $1 \leq i, k \leq \Psi(n)$, irrespective of their joint distribution and their dependence upon $\{d_{ik}\}$. We will, however, need to put more structure on $\{\theta_{ik}\}$ in Section V, where we focus our attention on an achievability scheme under LoS propagation. In this case, we have $\theta_{ik} = \frac{2\pi d_{ik}}{\lambda}$, where λ is the transmission wavelength. Further details about the LoS model are intentionally delayed until Section V, before which they are not needed.

We denote the network throughput, i.e., total number of bits/sec successfully decoded at the destinations in B_n , by $T(n)$. The *worst-case* achievable rate per source-destination pair (over all such pairs) is denoted by $R(n)$. The following probabilistic version of the ordering notation is used [7]. We write $f(n) = O(g(n))$ with high probability (w.h.p.) if \exists a constant K independent of n such that $\lim_{n \rightarrow \infty} \mathbb{P}(f(n) \leq Kg(n)) = 1$. Similarly, $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$.

In the same spirit, any general event \mathcal{A}_n is said to occur w.h.p. if $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{A}_n) = 1$. For notational simplicity, the bandwidth W is assumed to be 1 Hz and the noise power spectral density N_0 to be 1 watts/Hz. Also, each BS is assumed to have a maximum power constraint of P watts. Since $R(n)$ is defined as the worst-case rate, it is easy to establish that $T(n) = \Omega(nR(n))$ and hence $R(n) = O\left(\frac{T(n)}{n}\right)$.

III. INFORMATION THEORETIC UPPER BOUND

In this section, we derive an information-theoretic upper bound on the network throughput $T(n)$. We will make use of the following concentration Lemma.

Lemma 1. *For a homogeneous PPP $\Phi \subset \mathbb{R}^d$ with density λ_b , let $N(\mathcal{A})$ be the number of points in any measurable set $\mathcal{A} \subset \mathbb{R}^d$. Denote the Lebesgue measure of \mathcal{A} by $|\mathcal{A}|$. We have*

$$\lim_{|\mathcal{A}| \rightarrow \infty} \mathbb{P}(N(\mathcal{A}) \geq 2\lambda_b |\mathcal{A}|) = 0, \quad (2)$$

$$\lim_{|\mathcal{A}| \rightarrow \infty} \mathbb{P}\left(N(\mathcal{A}) \leq \frac{\lambda_b}{2} |\mathcal{A}|\right) = 0. \quad (3)$$

Proof: See Appendix A. ■

The upper bound is obtained through a cut-set bound argument. We partition the box B_n into two halves, each with side lengths $\sqrt{n} \times \sqrt{n}/2$, as shown in Fig. 1. We will study the information flow across the common edge of the two boxes, i.e., this edge acts as a cut. By Lemma 1, we have that w.h.p. there are less than $\lambda_b n$ points in each half-box. Since there are $O(n)$ source-destination pairs that need to transmit across this cut, the upper bound on the information flow across the cut also gives an upper bound (in order) for $T(n)$, from which the upper bound on source-destination rate $R(n)$ directly follows. Note that the information flow across the cut is upper bounded by the capacity of the effective MIMO channel, say C_n , with the BSs to the left of the cut operating as an effective transmitter and the BSs to the right as a receiver. Since we are interested in an upper bound, we can assume that there are exactly $\lambda_b n$ BSs on either side of the cut. Further, since the value of λ_b does not impact scaling results as long as it is finite, we fix it to 1 in this section for ease of notation. The $n\Psi(n) \times n\Psi(n)$ effective channel matrix is given by

$$\mathbf{H}_{\text{eff}} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \dots & \mathbf{H}_{1n} \\ \mathbf{H}_{21} & \mathbf{H}_{22} & \dots & \mathbf{H}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{n1} & \mathbf{H}_{n2} & \dots & \mathbf{H}_{nn} \end{bmatrix}, \quad (4)$$

where \mathbf{H}_{ik} is a $\Psi(n) \times \Psi(n)$ channel matrix from k^{th} BS from the left of the cut to the i^{th} BS to the right of the cut. Now, denoting the transmit symbol covariance matrix by \mathbf{Q} , we can write

$$\begin{aligned} C_n &= \max_{\substack{\mathbf{Q} \succeq 0 \\ \text{Tr}(\mathbf{Q}) \leq nP}} \log \det \left(\mathbf{I} + \mathbf{H}_{\text{eff}} \mathbf{Q} \mathbf{H}_{\text{eff}}^\dagger \right) \\ &\stackrel{(a)}{\leq} \log \det \left(\mathbf{I} + nP \mathbf{H}_{\text{eff}} \mathbf{H}_{\text{eff}}^\dagger \right) \end{aligned}$$

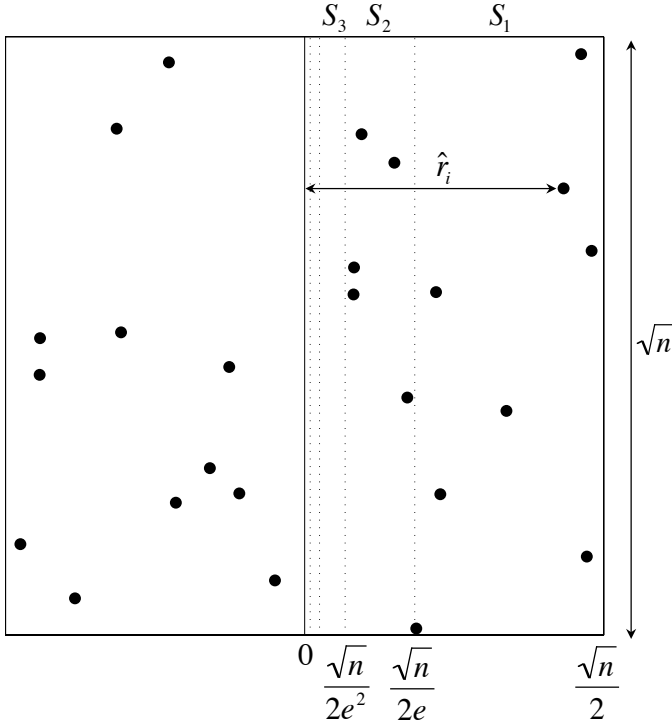


Fig. 1. The setup to derive information-theoretic upper bounds. The vertical strips are denoted by S_i , where $1 \leq i \leq \lfloor \log \frac{\sqrt{n}}{2} \rfloor + 1$.

$$\stackrel{(b)}{\leq} \sum_{i=1}^{n\Psi(n)} \log \left(1 + nP(\mathbf{H}_{\text{eff}}\mathbf{H}_{\text{eff}}^\dagger)_{ii} \right) \quad (5)$$

where (a) follows by the fact that $\log \det(\cdot)$ is monotonically increasing on the cone of positive semidefinite Hermitian matrices, and for any \mathbf{Q} satisfying $\text{Tr}(\mathbf{Q}) \leq nP$ we have $\mathbf{Q} \leq nP\mathbf{I}$,¹ and (b) follows from Hadamard inequality. For the notational ease, we assume that the distances between any pair of transmit and receive antennas between k^{th} BS on the left (transmitter) and the i^{th} BS on the right (receiver) are the same and equal to r_{ik} . The main idea now is to derive tight bounds on $(\mathbf{H}_{\text{eff}}\mathbf{H}_{\text{eff}}^\dagger)_{ii}$ using the geometric properties of the point process that determine the distances of BSs from the cut. The bound is based on the tools developed in [23, Theorem 5.4.4], [21], where a similar bound is derived for a single antenna network. A key difference is that the “mirroring argument” used in [21], [24] to establish equivalence between singular and eigenvalues of the channel matrix is not directly applicable in our case due to complex-valued channel gains. However, as discussed in the sequel, especially Appendix B, this requires only a few technical adjustments of the proof in [21].

We order the BSs on both sides of the cut by their respective distances from the cut. The distance of BS i from the cut is denoted by \hat{r}_i . To get a tight bound on $(\mathbf{H}_{\text{eff}}\mathbf{H}_{\text{eff}}^\dagger)_{ii}$, we use the *exponential stripping* technique introduced in [21]. As shown in Fig. 1, both the half-boxes on the either side of the cut are partitioned into $\lfloor \log \frac{\sqrt{n}}{2} \rfloor + 1$ vertical strips S_i .

For $1 \leq i \leq \lfloor \log \frac{\sqrt{n}}{2} \rfloor$, the minimum distance of the BSs lying in S_i from the cut is $\frac{\sqrt{n}}{2e^i}$, which will be used to upper bound $(\mathbf{H}_{\text{eff}}\mathbf{H}_{\text{eff}}^\dagger)_{ii}$. For $i = \lfloor \log \frac{\sqrt{n}}{2} \rfloor + 1$, i.e., the vertical strip closest to the cut, we will simply upper bound the path-loss by 1. Denoting the number of BSs in S_i by $X(S_i)$, the following holds w.h.p. $\forall i$

$$X(S_i) \leq \frac{n}{e^i} (e - 1). \quad (6)$$

The result for $1 \leq i \leq \lfloor \log \frac{\sqrt{n}}{2} \rfloor$ follows directly from Lemma 1 and for $i = \lfloor \log \frac{\sqrt{n}}{2} \rfloor + 1$ we have

$$\begin{aligned} & \lim_{|\mathcal{A}| \rightarrow \infty} \mathbb{P}(N(\mathcal{A}) > 2(e-1)\lambda|\mathcal{A}|) \\ & \leq \lim_{|\mathcal{A}| \rightarrow \infty} \mathbb{P}(N(\mathcal{A}) > 2\lambda|\mathcal{A}|) \stackrel{(a)}{=} 0, \end{aligned} \quad (7)$$

where (a) follows again from Lemma 1. From (5), we get

$$\begin{aligned} C_n & \leq \Psi(n) \sum_{i=1}^{\log \frac{\sqrt{n}}{2} + 1} X(S_i) \log \left(1 + nP(\mathbf{H}_{\text{eff}}\mathbf{H}_{\text{eff}}^\dagger)_{ii} \right) \\ & \leq (e-1) \Psi(n) \sum_{i=1}^{\log \frac{\sqrt{n}}{2} + 1} \frac{n}{e^i} \log \left(1 + nP(\mathbf{H}_{\text{eff}}\mathbf{H}_{\text{eff}}^\dagger)_{ii} \right), \end{aligned} \quad (8)$$

where we expressed $\lfloor \log \frac{\sqrt{n}}{2} \rfloor$ simply as $\log \frac{\sqrt{n}}{2}$ without compromising our results because we are interested in the behavior for $n \rightarrow \infty$. Recall that index i in $(\mathbf{H}_{\text{eff}}\mathbf{H}_{\text{eff}}^\dagger)_{ii}$ corresponds to all the BSs in the vertical strip S_i , which have the common upper bound

$$\begin{aligned} (\mathbf{H}_{\text{eff}}\mathbf{H}_{\text{eff}}^\dagger)_{ii} & = \Psi(n) \sum_{k=1}^n l(r_{ik}) \\ & \stackrel{(a)}{\leq} \Psi(n) \sum_{k=1}^n l(\hat{r}_i) = n\Psi(n)l(\hat{r}_i) \\ & \stackrel{(b)}{\leq} \begin{cases} n\Psi(n)\hat{r}_i^{-\alpha} & 1 \leq i \leq \log \frac{\sqrt{n}}{2} \\ n\Psi(n) & i = \log \frac{\sqrt{n}}{2} + 1 \end{cases} \\ & \stackrel{(c)}{\leq} \begin{cases} n^{1-\frac{\alpha}{2}}\Psi(n)2^\alpha e^{i\alpha} & 1 \leq i \leq \log \frac{\sqrt{n}}{2} \\ n\Psi(n) & i = \log \frac{\sqrt{n}}{2} + 1 \end{cases}, \end{aligned} \quad (10)$$

where (a) follows from the fact that $\hat{r}_i \leq r_{ik}$ and $l(\cdot)$ is a non-increasing function, (b) from the fact that for $1 \leq i \leq \log \frac{\sqrt{n}}{2}$, \hat{r}_i , which is the distance of the BS lying in S_i from the cut, is lower bounded by $\frac{\sqrt{n}}{2e^i} \geq 1$, which implies $l(\hat{r}_i) = \hat{r}_i^{-\alpha}$, and for $i = \log \frac{\sqrt{n}}{2} + 1$, we simply upper bound $l(\hat{r}_i)$ by 1, and (c) follows by lower bounding \hat{r}_i . Substituting (10) in (9)

$$\begin{aligned} C_n & \leq (e-1) \Psi(n) \sum_{i=1}^{\log \frac{\sqrt{n}}{2}} \frac{n}{e^i} \log \left(1 + Pn^{2-\frac{\alpha}{2}}\Psi(n)2^\alpha e^{i\alpha} \right) \\ & \quad + 2\frac{e-1}{e} \Psi(n) \sqrt{n} \log \left(1 + Pn^2\Psi(n) \right), \end{aligned} \quad (11)$$

where the last term is $O(\Psi(n)\sqrt{n} \log(n^2\Psi(n)))$. Denoting the summation in the first equation by C_s , we can express (11) as

$$C_n \leq (e-1) \Psi(n) C_s + O(\Psi(n)\sqrt{n} \log(n^2\Psi(n))). \quad (12)$$

¹For matrices \mathbf{A} and \mathbf{B} we write $\mathbf{A} \leq \mathbf{B}$ to indicate that the difference $\mathbf{B} - \mathbf{A}$ is positive semi-definite.

After some effort, we can prove the following result on the scaling of C_s .

Lemma 2. For path-loss exponent $\alpha > 2(2 + \log_n \Psi(n))$

$$C_s = O\left(\sqrt{n} n^{\frac{2}{\alpha}} \Psi(n)^{\frac{1}{\alpha}} \log n\right). \quad (13)$$

Proof: See Appendix B. ■

From Lemma 2, it is clear that the first term of (12) scales as $O\left(\sqrt{n} n^{\frac{2}{\alpha}} \Psi(n)^{1+\frac{1}{\alpha}} \log n\right)$, which represents the dominating term. This leads to the following upper bound on $T(n)$.

Theorem 1. For path-loss exponent $\alpha > 2(2 + \log_n \Psi(n))$

$$T(n) = O\left(\sqrt{n} n^{\frac{2}{\alpha}} \Psi(n)^{1+\frac{1}{\alpha}} \log n\right). \quad (14)$$

Corollary 1. The number of antennas needed per-BS in order to achieve $R(n) = O(1)$ for $\alpha > 4$ is

$$\Psi(n) = \Omega\left(\left[n^{\frac{1}{2}-\frac{2}{\alpha}} (\log n)^{-1}\right]^{\frac{\alpha}{1+\alpha}}\right). \quad (15)$$

Remark 1 (Scalability). The above Corollary should be interpreted as a lower bound on the number of antennas per BS needed for scalability. In particular, for high attenuation regime, we need to scale antennas almost as \sqrt{n} to make the backhaul network scalable. In fact, in Section V, $\Psi(n) = \sqrt{n}$ is shown to achieve scalability for any $\alpha > 2$ in LoS MIMO networks, thus showing that there exists a regime of BS physical sizes, inter-BS distances, and high frequency, where a scalable wireless backhaul can be effectively implemented with short hops, each achieving high MIMO multiplexing gain.

IV. LONG HOPS: BEAMFORMING

In this section, we consider a transmission strategy where each BS uses all its antennas for beamforming, i.e., it transmits a single data stream to the farthest possible BS in the direction of its destination in order to minimize the number of hops. Our goal is to find an upper bound on the achievable $R(n)$ under this strategy, as a function of $\Psi(n)$ and α . Recall that for a given link with transmitter-receiver separation of $d > 1$, and $\Psi(n) \times \Psi(n)$ channel matrix \mathbf{H} with entries given by (1), the maximum rate achievable for a single stream under eigen-beamforming is

$$\begin{aligned} \text{Rate}(d) &\stackrel{(a)}{=} \log(1 + P\lambda_{\max}(\mathbf{H}\mathbf{H}^\dagger)) \\ &\stackrel{(b)}{\leq} \log(1 + Pd^{-\alpha}\Psi(n)^2), \end{aligned} \quad (16)$$

where $\lambda_{\max}(\mathbf{H}\mathbf{H}^\dagger)$ in (a) is the maximum eigenvalue of $\mathbf{H}\mathbf{H}^\dagger$, and (b) follows from the fact that $\lambda_{\max}(\mathbf{H}\mathbf{H}^\dagger) \leq \text{Tr}(\mathbf{H}\mathbf{H}^\dagger) = d^{-\alpha}\Psi(n)^2$. To minimize the number of hops for a given source-destination pair, our goal is to maximize distance d for each hop keeping the transmission rate (16) constant. Assuming a minimum target received power equal to some value P_0 , we have

$$Pd^{-\alpha}\Psi(n)^2 \geq P_0 \Rightarrow d \leq \left(\frac{P}{P_0}\right)^{\frac{1}{\alpha}} \Psi(n)^{\frac{2}{\alpha}} = d_c. \quad (17)$$

Given d_c , we need a lower bound on the source-destination separation in order to lower bound the number of hops needed. This is provided by the following Lemma.

Lemma 3 (Lower bound on source-destination separation). The source-destination separation of a randomly chosen pair in B_n is $\Omega(n^{\frac{1}{2}-\epsilon})$ w.h.p., where $\epsilon > 0$.

Proof: See Appendix C. ■

Then, the number of hops N_h required for a randomly chosen source-destination pair is lower bounded as $N_h = \Omega\left(\frac{n^{\frac{1}{2}-\epsilon}}{d_c}\right)$. From Lemma 1, we have that the number of BSs in B_n is $\Omega(n)$ and therefore the number of source-destination pairs is also $\Omega(n)$. As a result, a lower bound on the total number of hops needed in the network is $nN_h = \Omega\left(\frac{n^{\frac{3}{2}-\epsilon}}{d_c}\right)$. Recall that each hop achieves a fixed target receiver power P_0 , therefore it can support a fixed peak rate, which needs to be shared among all the connections handled by the “bottleneck” BS. Hence, there is at least one BS that has to relay $\Omega\left(\frac{n^{\frac{1}{2}-\epsilon}}{d_c}\right)$ connections. Therefore, the rate per source-destination pair is upper bounded by $R(n) = O\left(d_c n^{-\frac{1}{2}+\epsilon}\right)$, which leads to the following main result.

Theorem 2 (Long hop). For the long hop strategy discussed in this section

$$R(n) = O\left(\frac{\Psi(n)^{\frac{2}{\alpha}}}{n^{\frac{1}{2}-\epsilon}}\right). \quad (18)$$

Remark 2 (Scalability under long-hop strategy). From Theorem 2, it is clear that in order to achieve $R(n) = O(1)$, we need $\Psi(n) = \Omega(n^{\frac{\alpha}{4}-\epsilon})$, which for a vanishingly small ϵ and $\alpha > 2$ is always higher than \sqrt{n} . In fact, for $\alpha = 4$, we need to scale the number of antennas almost linearly with n .

V. SHORT HOPS IN LOS: SPATIAL MULTIPLEXING

In this section, we consider the other extreme where the data for each source-destination pair is communicated through a series of *short hops*, also termed as *information highways* in the context of random networks [7]. For the ease of exposition, we introduce the key ideas using a simpler, but still meaningful, model shown in Fig. 2, where the BS locations are given by a perturbed square lattice. The distance between the closest lattice points is assumed to be a constant $c = \lambda_b^{-\frac{1}{2}}$ in order to preserve the BS density λ_b as in Section II. After establishing the results for this simpler model, we will generalize them to the random network using percolation theory arguments of [7]. By choosing this simpler model for exposition, we avoid repeating many key arguments of [7], e.g., the construction of bond percolation model, while still conveying the main message.

By short hop, we specifically refer to the communication link between two neighboring BSs, each lying in adjacent small squares in Fig. 2. As reviewed in Section I-A, in classical single-antenna wireless networks the short-hop strategy achieves $R(n) = \Omega\left(\frac{1}{\sqrt{n}}\right)$. The main intuition behind this result can be explained in terms of *horizontal* and *vertical* routes, formed by rows and columns of small squares in Fig. 2, running from left to right or from top to bottom edges of B_n . Note that under single-antenna transmission, both the horizontal and vertical routes achieve rate $\Omega(1)$ if the interference

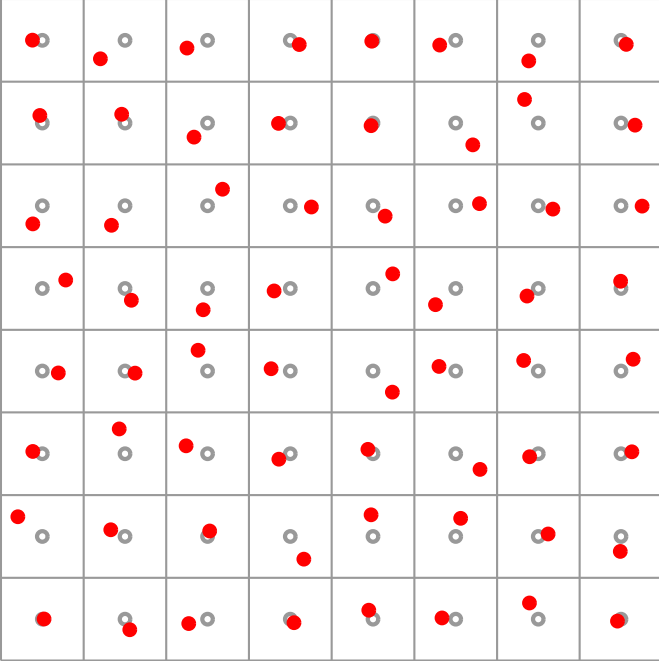


Fig. 2. Perturbed lattice model for the BS locations. The lattice and the BS locations are denoted by hollow and filled circles, respectively.

power at each BS is bounded. The result now follows from the fact that any destination can be reached from its source by following a simple routing strategy, where data is first sent over a horizontal route until it reaches the column where its destination lies, after which it is sent over that vertical route. In other words, each route is shared by \sqrt{n} connections, which implies that the rate per connection is $R(n) = \Omega\left(\frac{1}{\sqrt{n}}\right)$. A natural question to ask now is what happens when these short hops are MIMO links capable of transmitting multiple independent streams by spatial multiplexing. If the number of streams remains constant independent of n , each route achieves a higher but constant rate, which does not affect the scaling results. However, if they scale up as $\Omega(\sqrt{n})$, the rate of each route scales as \sqrt{n} , which implies a per source-destination rate of $R(n) = \Omega(1)$. Therefore, the main goal of this section is to find $\Psi(n)$ which enables each short hop MIMO link to achieve rate, say $R_{\text{sh}}(n)$, of $\Omega(\sqrt{n})$ in the presence of interference due to other simultaneous MIMO transmissions.

Since this section deals with an achievability result, we need to be careful with the channel model in order to avoid conflicting conclusions. We explicitly consider LoS propagation due to the following reasons: (i) LoS propagation model is consistent with our general channel gain definition given by (1); (ii) it is relevant in the context of high carrier frequencies (mm-waves range) for which “rich scattering” has not been observed in channel measurements. In fact, at such high frequencies, the rank of non-LoS channel matrices is limited by the number of dominant scatterers rather than by the number of antennas (see the recent work in [10] and references therein); (iii) LoS propagation for the backhaul can be obtained by roof-top mounted antenna arrays, especially in sub-urban

areas where homes have roughly the same height; (iv) the structure of LoS channel matrices yield a natural collapse of channel DoFs as a function of geometry, thereby avoiding the annoying dichotomy of results that occurs under the rich scattering assumption (discussed in Section I-A). Before going into more technical details, we review the key results for LoS propagation in the context of our model.

A. LoS Propagation

We assume that the antennas at each BS node are uniformly distributed in squares of side \sqrt{a} , significantly smaller than the inter-BS distance. In practice, for finite number of antennas, the orientation of arrays and the placement of antenna elements can be carefully optimized to maximize the rate achievable by a LoS MIMO link, e.g., see [25]. For $\Psi(n) \rightarrow \infty$, the achievable spatial DoFs for a LoS $\Psi(n) \times \Psi(n)$ MIMO channel were independently derived in [14], [15] to be

$$\begin{cases} \min \left\{ \Psi(n), \frac{\sqrt{a}}{\lambda} \right\}, & \text{when } d \in [1, \sqrt{a}] \\ \min \left\{ \Psi(n), \frac{a}{\lambda d} \right\}, & \text{when } d \in \left(\sqrt{a}, \frac{a}{\lambda} \right] \\ 1, & \text{when } d \in \left(\frac{a}{\lambda}, \infty \right), \end{cases} \quad (19)$$

where λ denotes the carrier wavelength and d is the transmitter-receiver distance. A matching upper bound (within logarithmic factors) on the DoFs has been recently derived in [26] under an element-by-element approximation of the LoS MIMO channel matrix by another random matrix. As discussed in the sequel, a particular case of interest for short hop strategy is when a LoS MIMO link can achieve DoFs $= \Psi(n)$ for all $d \leq d_{\text{max}}$, for some constant d_{max} independent of n . From (19), it is clear that for $d_{\text{max}} \leq \sqrt{a}$ this can be achieved by: (i) fixing λ and scaling $\sqrt{a} \propto \Psi(n)$, or (ii) fixing a and scaling $\lambda \propto \Psi(n)^{-1}$. In a realistic urban backhaul network, d_{max} would correspond to the separation between closest BSs and would be of the order of 100s of meters. Therefore, for reasonable form factors of the BSs, $d_{\text{max}} \leq \sqrt{a}$ will never hold in practice. Now, focusing on the more realistic case of $d_{\text{max}} > \sqrt{a}$, DoFs $= \Psi(n)$ can be achieved by: (i) fixing λ and scaling $a \propto \Psi(n)$, or (ii) fixing a and scaling $\lambda \propto \Psi(n)^{-1}$. While it is not possible to keep growing the size of BS with $\Psi(n)$, it is indeed possible to assume $\lambda \propto \Psi(n)^{-1}$, which is consistent with the ongoing migration towards higher transmission frequencies [9]. An example of a possible practical layout is given next.

Remark 3 (Sub-urban small-cell network). *Consider an urban wireless backhaul network with $d_{\text{max}} = 100\text{m}$. Also consider $\Psi(n) = 64$, which we know is possible with the current technology [27]. To achieve full DoFs at a carrier frequency of 30GHz, i.e., $\lambda = 1\text{cm}$, we need $a = \Psi(n)\lambda d_{\text{max}} = 64\text{m}^2$, which means a square array of side length 8m. This yields a quite practical setup, where $\Psi(n)$ radiating elements for each BS can be integrated into roof or other architectural elements in order to enable a high-throughput wireless backhaul connectivity. Notice also that these antennas may be for backhaul-only purposes, since the cellular uplink and downlink can be handled at different frequencies through a different geometry (involving both BSs and users).*

The assumption $\lambda \propto \Psi(n)^{-1}$ requires a careful treatment of link budget and pathloss, which we do next. In LoS conditions, the channel gain between two single-antenna nodes with separation $d \gg \lambda$ is [14]

$$h = \sqrt{G_T G_R} \frac{\lambda}{4\pi d} \exp\left(-j \frac{2\pi}{\lambda} d\right) \stackrel{(a)}{=} \frac{\sqrt{G_T G_R}}{4\pi d \Psi(n)} \exp(-j 2\pi d \Psi(n)), \quad (20)$$

where G_T and G_R are the antenna gains at the transmitter and the receiver, respectively, and (a) follows from $\lambda \propto \Psi(n)^{-1}$ with proportionality constant of 1 (choice of constant does not matter for the sake of scaling laws). In the context of our model, G_T and G_R refer to the gains of a single radiating element, not to the array gain of the $\Psi(n)$ -element array. As discussed in [28, Section 4.10], for a radiating element of physical area A , the antenna gain is $G \propto \frac{A}{\lambda^2}$. Scaling up the number of array elements as $\Psi(n)$, while keeping the total size of the array constant, implies $A \propto \Psi(n)^{-1}$. This along with the fact that $\lambda \propto \Psi(n)^{-1}$ implies that both G_T and G_R increase linearly with $\Psi(n)$. Substituting this back in (20) for both G_T and G_R (within an irrelevant proportionality constant which is taken equal to 1 for simplicity of notation), we get

$$h = (4\pi d)^{-1} \exp(-j 2\pi d \Psi(n)). \quad (21)$$

Note that, although uncommon, the scaling of frequency with n has been considered before [14]. In this section, we adopt a slightly modified version of (21). In particular, the gain between the k^{th} antenna of the transmitter and the i^{th} antenna of the receiver is

$$h_{ik} = \min\{d_{ik}^{-\frac{\alpha}{2}}, 1\} \exp(-j 2\pi d_{ik} \Psi(n)), \quad (22)$$

where the pathloss exponent $\alpha > 2$ captures larger attenuation that may result from roof-top diffraction [15], and $\min\{d_{ik}^{-\frac{\alpha}{2}}, 1\}$ avoids singularity at the origin. The constant $\frac{1}{4\pi}$ is irrelevant for the scaling results and it is ignored in (22). For notational simplicity, we assume amplitude term $\min\{d_{ik}^{-\frac{\alpha}{2}}, 1\}$ in (22) to be the same and equal to $\min\{d^{-\frac{\alpha}{2}}, 1\}$ for all the transmit-receive antenna pairs for a given link due to the assumption that the BS linear size \sqrt{a} is significantly smaller than the separation between two BSs (see Remark 3). The LoS model (22) is consistent with the general model given by (1).

B. Throughput for the Perturbed Lattice Model

The main challenge in analyzing $R_{sh}(n)$ is the presence of LoS interference originating from other simultaneous MIMO transmissions. A closely related problem has recently been studied in the literature as a part of the *hierarchical cooperation strategy* for ad hoc networks, where one of the intermediate steps is to derive the rate achievable by a distributed MIMO transmission in the presence of interference from other simultaneous MIMO transmissions [12]. While the procedure to handle this interference under i.i.d. MIMO channels is well understood, it is not the case when the interfering MIMO links are LoS [14], [15]. This problem is rigorously treated in [14], where the distributed MIMO link rate is derived by explicitly considering LoS MIMO interfering links, see [14, Lemma 2].

In this section, we take an alternate route and study the scaling of $\mathbb{E}[R_{sh}(n)]$, where expectation is over the (random) antenna locations at each BS. We show that $\Psi(n) = \sqrt{n}$ is sufficient to achieve $\mathbb{E}[R_{sh}(n)] = \Omega(\sqrt{n})$. Our analysis is considerably simpler and involves a direct bound on the interference power, reducing the problem to finding the spatial DoFs of a single LoS MIMO link in isolation, which is given by (19). We then remark on the connections of this result with the scaling of $R_{sh}(n)$. We begin with the following Lemma.

Lemma 4. For any continuous random variable $D \geq 0$

$$\lim_{\Psi(n) \rightarrow \infty} \mathbb{E}[\exp(-j 2\pi D \Psi(n))] = 0. \quad (23)$$

Proof: See Appendix D. ■

Assume $c\sqrt{5} \leq d_{\max}$ so that each short-hop achieves full spatial DoFs in the absence of interference (recall that c is the distance between adjacent points in the lattice of Fig. 2). Denote the LoS MIMO channel matrix of the desired link by \mathbf{H} and that of the i^{th} interferer to the desired receiver by $\mathbf{H}^{(i)}$. Similarly, denote the transmit symbol of the desired and i^{th} interfering transmitters by \mathbf{x} and $\mathbf{x}^{(i)}$, respectively. We further assume that each transmitter distributes equal power across antennas, i.e., $\mathbb{E}[\mathbf{x}\mathbf{x}^\dagger] = \mathbb{E}[\mathbf{x}^{(i)}\mathbf{x}^{(i)\dagger}] = \frac{P}{\Psi(n)}\mathbf{I}$. For worst case analysis, we assume all the nodes are transmitting.² Denoting the set of interferers by \mathcal{I} , and the noise vector by \mathbf{z} , the received signal at the desired receiver is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \sum_{i \in \mathcal{I}} \mathbf{H}^{(i)}\mathbf{x}^{(i)} + \mathbf{z}. \quad (24)$$

The rate achievable by a short-hop link can be lower-bounded as

$$R_{sh}(n) \geq \log \det \left(\mathbf{I} + \frac{P}{\Psi(n)} \mathbf{R}^{-1} \mathbf{H} \mathbf{H}^\dagger \right), \quad (25)$$

where \mathbf{R} is the covariance matrix of noise-plus-interference observed at the desired receiver, given by

$$\mathbf{R} = \mathbf{I} + \frac{P}{\Psi(n)} \sum_{i \in \mathcal{I}} \mathbf{H}^{(i)} \mathbf{H}^{(i)\dagger}. \quad (26)$$

Since the antenna locations of the interfering transmitters only affect (25) through \mathbf{R} , by applying iterated expectation and Jensen's inequality we have

$$\mathbb{E}[R_{sh}(n)] \geq \mathbb{E} \left[\log \det \left(\mathbf{I} + \frac{P}{\Psi(n)} \mathbb{E}[\mathbf{R}]^{-1} \mathbf{H} \mathbf{H}^\dagger \right) \right], \quad (27)$$

where, with a slight abuse of notation, the inner expectation is with respect to the antenna locations of the interfering transmitters conditioned on the antenna locations of the receiver, and the outer expectation is with respect to the antenna locations of the receiver and the intended transmitter. Now the goal is to upper bound the inner (conditional) expectation $\mathbb{E}[\mathbf{R}]$, which is

$$\mathbb{E}[\mathbf{R}] = \mathbf{I} + \frac{P}{\Psi(n)} \sum_{i \in \mathcal{I}} \mathbb{E} \left[\mathbf{H}^{(i)} \mathbf{H}^{(i)\dagger} \right], \quad (28)$$

²We could have used a reuse factor larger than 1 to control interference temperature, e.g., see [12], but this does not affect our scaling results because of the bounded pathloss model.

where the $(k, m)^{th}$ element of $\mathbf{H}^{(i)}\mathbf{H}^{(i)\dagger}$ is

$$\sum_{l=1}^{\Psi(n)} \left[\min \left\{ 1, \left(d^{(i)} \right)^{-\frac{\alpha}{2}} \right\} \right]^2 e^{-j2\pi d_{kl}^{(i)} \Psi(n)} e^{j2\pi d_{ml}^{(i)} \Psi(n)}. \quad (29)$$

Taking expectation in (29) with respect to the antenna locations of the interfering transmitters and using Lemma 4, we have

$$\mathbb{E}[\mathbf{R}] \rightarrow \left(1 + \sum_{i \in \mathcal{I}} P \left[\min \left\{ 1, \left(d^{(i)} \right)^{-\frac{\alpha}{2}} \right\} \right]^2 \right) \mathbf{I} \quad (30)$$

for large $\Psi(n)$. Note that there are at most 8 squares in the closest *ring* of interferers around the square in which the desired receiver is located, such that the total interference power contributed by this first ring of interferers to the summation term of (30) is upper bounded by $8P$. Similarly, for $i > 1$, there are at most $8i$ interferers in the i^{th} ring, each at least at a distance $(i-1)c$, which upper bounds the interference contribution from the i^{th} ring by $8i(c(i-1))^{-\alpha}$. This counting argument leads to the following upper bound on $\mathbb{E}[\mathbf{R}]$

$$\mathbb{E}[\mathbf{R}] \leq \left(1 + 8P + 8c^{-\alpha} P \sum_{i=2}^{\infty} i(i-1)^{-\alpha} \right) \mathbf{I}, \quad (31)$$

that holds for sufficiently large $\Psi(n)$, where the summation $\sum_{i=2}^{\infty} i(i-1)^{-\alpha} \leq \sum_{i=2}^{\infty} i^{1-\alpha}$ is convergent for $\alpha > 2$. Therefore, in the asymptotic regime of large $\Psi(n)$, have that $\mathbb{E}[\mathbf{R}] \leq q\mathbf{I}$, for some constant q independent of n . Substituting it back in (25), we get

$$\mathbb{E}[R_{sh}(n)] \geq \mathbb{E} \left[\log \det \left(\mathbf{I} + \frac{P}{q\Psi(n)} \mathbf{H}\mathbf{H}^\dagger \right) \right] \stackrel{(a)}{=} \Omega(\Psi(n)), \quad (32)$$

where (a) follows from the spatial DoFs result given by (19). This leads to our second main result.

Theorem 3 (Short hop: perturbed lattice). *For the short hop strategy and the perturbed lattice model for the BS locations, $\Psi(n) = \sqrt{n}$ achieves $\mathbb{E}[R_{sh}(n)] = \Omega(\sqrt{n})$ for each short hop, and hence ergodic rate of $\Omega(1)$ for each source-destination pair; where expectation is over antenna locations.* ■

Remark 4 (Scalability). *Comparing Theorems 2 and 3, we note that short hop strategy is significantly better for network scalability. While for any $\alpha > 2$ it requires only $\Psi(n) = \sqrt{n}$, the antenna requirement in long hop is always higher and keeps increasing further with α .*

Note that Theorem 3 would be enough to claim achievability of the scaling law if the BSs were allowed to randomly select their antenna locations independently and uniformly in their allowed fixed area, such that the expectation of (25) has the operational meaning of “ergodic rate”, achieved by coding over a long sequence of realizations of the antenna locations. A more practical viewpoint is to consider the achievability of the scaling law with high probability for a fixed random realization of the antennas. For such a result, we need to show that $\mathbb{P}(R_{sh}(n) \geq r\Psi(n)) \rightarrow 1$ as $n \rightarrow \infty$, for some constant $r > 0$. One way to establish this result is by showing that $R_{sh}(n)$ concentrates around its mean, in the sense that

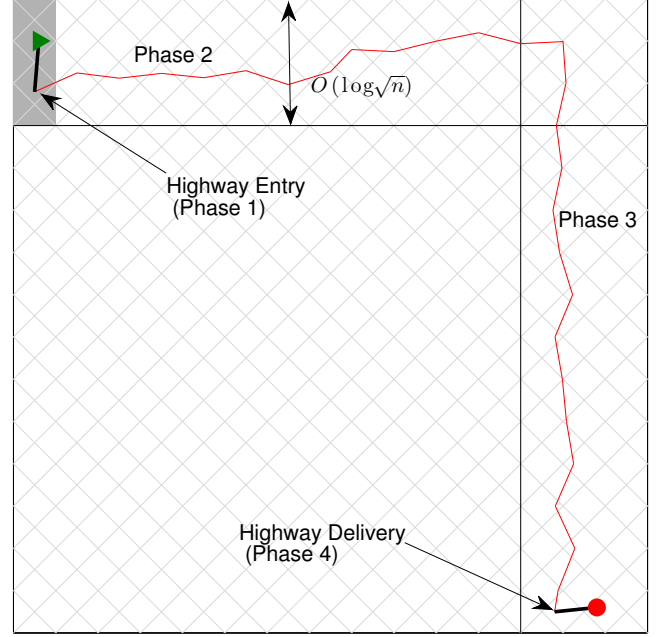


Fig. 3. A schematic sketch of 4-phase routing strategy for a random network. The highway entry (phase 1) and exit (phase 4) are denoted by thicker lines. The triangle and circle denote source and destination BSs, respectively.

$\mathbb{P}(|R_{sh}(n) - \mathbb{E}[R_{sh}(n)]| > \epsilon\Psi(n)) \rightarrow 0$, with $\epsilon > 0$, as $n \rightarrow \infty$. Such concentration result is shown in [15, Lemma 2.2] for the case of an isolated (without interference) LoS MIMO link. In Section VI, we provide numerical evidence of this concentration. Complementary to this and the achievability proof in [14, Lemma 2], we propose a simpler approach in Appendix E based on the spectral radius results of [29]. Note that Appendix E merely suggests a direction for future work and should not be interpreted as a formal result due to two reasons: (i) we use a slightly tighter bound on spectral radius than [29] to illustrate the approach (the tighter bound needs proof), and (ii) the bound in [29] is itself derived under an element-wise approximation of the LoS channel matrix by another random matrix, which needs to be rigorized.

C. Throughput for the Random Network Model

It is easy to generalize the above achievability result to the random network model using percolation theory [7]. The key ideas are explained with the help of Fig. 3. The source and the destination BSs of a given pair are denoted by a triangle and circle, respectively. The routing protocol is formed by four phases. In the first phase, the source BS sends its data to a *horizontal highway*. In the second phase, the data is communicated over a horizontal highway, until it reaches the *vertical highway* associated with the destination BS. In the third phase, data is transmitted over this vertical highway. Finally, in the fourth phase, the data is delivered to the destination by the last hop off the highway.

In [7], it is shown that w.h.p. there are $\Omega(\sqrt{n})$ *horizontal highways*, connecting the left and right edges of B_n such that the distance between any two neighboring BSs on each highway is bounded by a constant independent of n . Similarly

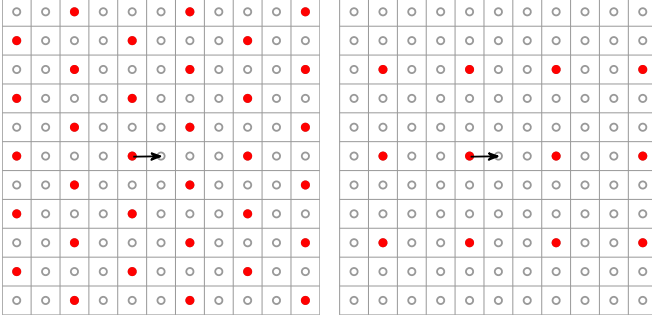


Fig. 4. Deterministic reuse patterns: (first) 4-reuse, and (second) 9-reuse. The BSs active in the current time slot are denoted by solid circles. The LoS MIMO link under investigation is denoted by a one-sided arrow.

there are $\Omega(\sqrt{n})$ highways connecting the top and bottom edges. It is further shown in [7] that these highways are distributed uniformly over the network. In particular, these horizontal and vertical highways can be respectively grouped into disjoint groups of $\log(\sqrt{n})$ highways, with each group confined in a rectangular slab of size $\sqrt{n} \times \log \sqrt{n}$. One each of such horizontal and vertical slabs is also illustrated in Fig. 3. Due to this regularity property, it is shown in [7] that it is possible to uniquely associate a slab of width $O(1)$ with each highway, such that each highway carries data of $O(\sqrt{n})$ source-destination pairs. Hence, phases 2 and 3 of the above routing protocol are equivalent to what was already discussed in the context of perturbed lattice model. In particular, with $\Psi(n) = \sqrt{n}$ it is possible for each highway to simultaneously achieve rate $\Omega(\sqrt{n})$, where it should be noted that the interference power can be bounded in exactly the same way as demonstrated in the proof of Theorem 3. This rate is then equally distributed over $O(\sqrt{n})$ connections, such that each achieves rate $\Omega(1)$, which leads to the above discussed properties. Interested readers can refer to [7] for more details about this construction.

Now, we need to show that phases 1 and 4 are not the rate bottlenecks. The argument is the same for both the phases, so we explain it in terms of phase 1. The width of the horizontal and vertical slabs yields that the distance between a source and the nearest highway node is $O(\log \sqrt{n})$. Substituting this distance in (19) and recalling that $\lambda \propto \Psi(n)^{-1}$, it is easy to see that the spatial DoFs for this channel are $\Omega\left(\frac{\sqrt{n}}{\log \sqrt{n}}\right)$ and that the interference power can be bounded in the same way as done in the previous subsection. To complete the argument, we have to find the number of BSs sharing the same highway entry point. The rate $\Omega\left(\frac{\sqrt{n}}{\log \sqrt{n}}\right)$ must be shared amongst these BSs. As shown in Fig. 3, these source BSs are confined in the gray region of size $O(1) \times O(\log \sqrt{n})$. By Lemma 1, the number of BSs in this region is upper bounded by $O(\log \sqrt{n})$. Therefore, the rate per source node in phase 1 is lower bounded by $\Omega\left(\frac{\sqrt{n}}{(\log \sqrt{n})^2}\right)$, which shows that phases 1 and 4 are not the rate bottleneck of the system.

VI. DISCUSSION

In this section we discuss some additional aspects and consequences of throughput scaling analysis of wireless backhaul

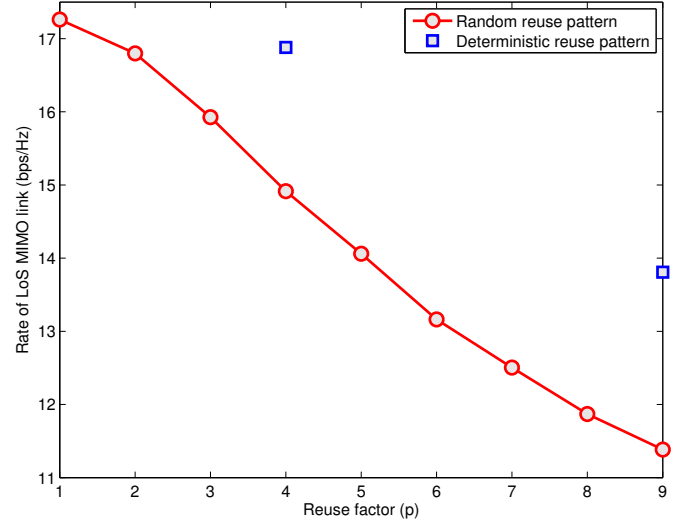


Fig. 5. Rate of LoS MIMO link as a function of the reuse factor. In addition to the random reuse, the deterministic reuse patterns of Fig. 4 are also simulated.

networks carried out in the previous sections.

a) *Numerical Evaluation of Achievable Rates:* In Section V we showed that it is possible to implement a scalable wireless backhaul network using short hop strategy, even in LoS propagation. In this section, we take a step forward and find a ball-park number for the rate achievable by a short hop link. For simplicity, we assume that the BSs are located on a squared lattice. To minimize edge effects, we simulate a lattice with 529 BSs, with the desired receiver at the lattice center (see Fig. 4 where the short hop link of interest is denoted by a one-sided arrow). We assume carrier frequency of 30GHz and pathloss exponent $\alpha = 5$.³ Letting $\mu = \frac{P_c^{-\alpha}}{N_0 W}$ denote the SNR for the desired link under single antenna transmission in the absence of interference, we set the desired reference value $\mu = 0$ dB, which means that such single antenna link in isolation would achieve a rate of $\log_2(1 + \mu) = 1$ bps/Hz. For the LoS MIMO link, we use the same parameters as in Remark 3, with inter-BS distance of 100m, and $\Psi(n) = 64$ BS antennas uniformly distributed in a square of side 8m. The achievable rate for the desired link can be numerically computed using (25). Assuming that all the BSs transmit simultaneously (i.e., *full-reuse*), the rate is about 17 bps/Hz, which means that with only about 100 MHz bandwidth (well within what it is anticipated in the 30 GHz band), a 1.7 Gbps backhaul link can be implemented with fairly standard technology.

b) *Effect of Time/Frequency Reuse:* As is usually done in cellular systems, it is possible to introduce a resource reuse factor $p > 1$ in a wireless backhaul system as well, i.e., the network is partitioned into p sub-networks, each of which is active on a fraction $1/p$ of the total transmission resource (time-frequency slots). Fig. 4 shows two examples for a square lattice network with $p = 4$ and $p = 9$ reuse factors. While it was possible to partition the BSs into p meaningful sets

³Note that the characterization of pathloss for mm-wave communication is currently under active investigation (see [8] for some recent measurement results).

in these two cases, it is not always easy to accomplish this for any given p . Therefore, we also consider a *random reuse* strategy where we randomly select with uniform probability $1/p$ the BSs using the same resource. In all cases, the link rate with reuse p is given by

$$R_{\text{sh}}(n) \geq \frac{1}{p} \log \det \left(\mathbf{I} + \frac{pP}{\Psi(n)} \mathbf{R}^{-1} \mathbf{H} \mathbf{H}^\dagger \right), \quad (33)$$

where \mathbf{R} now includes interference from only those BSs that are active on the same resource partition as the desired link, and the transmit power is increased by the factor p since each BS is active only on a fraction $1/p$ of the channel slots. For the same simulation setup as the previous subsection, we numerically evaluate (33) for various values of p and plot the mean rates in Fig. 5. Two remarks are in order: (i) for random reuse, $p = 1$ (universal reuse) is the best option for this set of simulation parameters; and (ii) deterministic (planned) reuse performs significantly better both for $p = 4$ and 9.

c) *Distribution of the rate of LoS MIMO Link*: While the previous two subsections focused on the mean rates achievable by a LoS MIMO link, in this subsection we explore the link rate distribution. We consider the same setup as the previous two subsections and evaluate the achievable rate using (25) with full reuse. Instead of averaging with respect to the antenna placement, we plot the rate distributions in Fig. 6 for two cases: (i) $\Psi(n) = 64$, and (ii) $\Psi(n) = 256$. We also plot the mean rate obtained by Monte Carlo simulation and the “ergodic” lower bound on the mean rate given by (27). Note that the rate distribution is fairly concentrated around its mean, and this concentration is more and more evident when the number of antennas increases. Note also that the ergodic rate lower bound developed and analyzed in this paper is concisely on the left of the actual rate distribution. As discussed in Section V-B after Remark 4, this indicates that the actual achievable rate for a random placement of the BS antennas has the same scaling as the ergodic lower bound w.h.p. (a stronger achievability conclusion than Theorem 3). See also Appendix E for more discussion on this point.

d) *Availability and Impact of Wired Backhaul*: Throughout this paper, we assumed that none of the BSs has access to wired backhaul. This can be easily relaxed in some cases of practical relevance. For instance, let B_n represent an urban cellular network with some of the BSs (gateways) at the boundary (perimeter) having access to wired backhaul. It is reasonable to assume that each cell within B_n generates fraction ρ of traffic with destinations inside B_n and a fraction $1 - \rho$ of traffic with destinations outside the network. Therefore, for any $\rho \in (0, 1)$, $\Theta(n)$ traffic must flow through the gateways into the wired backbone network. Following short hop strategy with $\Psi(n) = \sqrt{n}$, each gateway BS can relay $\Omega(\sqrt{n})$ traffic to the “outside world”, which means we need $O(\sqrt{n})$ gateways at the boundary of B_n to relay a total of $\Theta(n)$ traffic.

Another setup of interest is when $O(n^\beta)$, $\beta \in [0, 1]$ BSs inside B_n have access to high-capacity wired backhaul. For the ease of exposition, let us consider the lattice model with regular placement of the wired backhaul BSs. Then, this setup reduces to n^β mini-networks of $n^{1-\beta}$ BSs each. By Theorem 3, each BS needs $\Psi(n) = \Omega(n^{\frac{1-\beta}{2}})$ antennas to

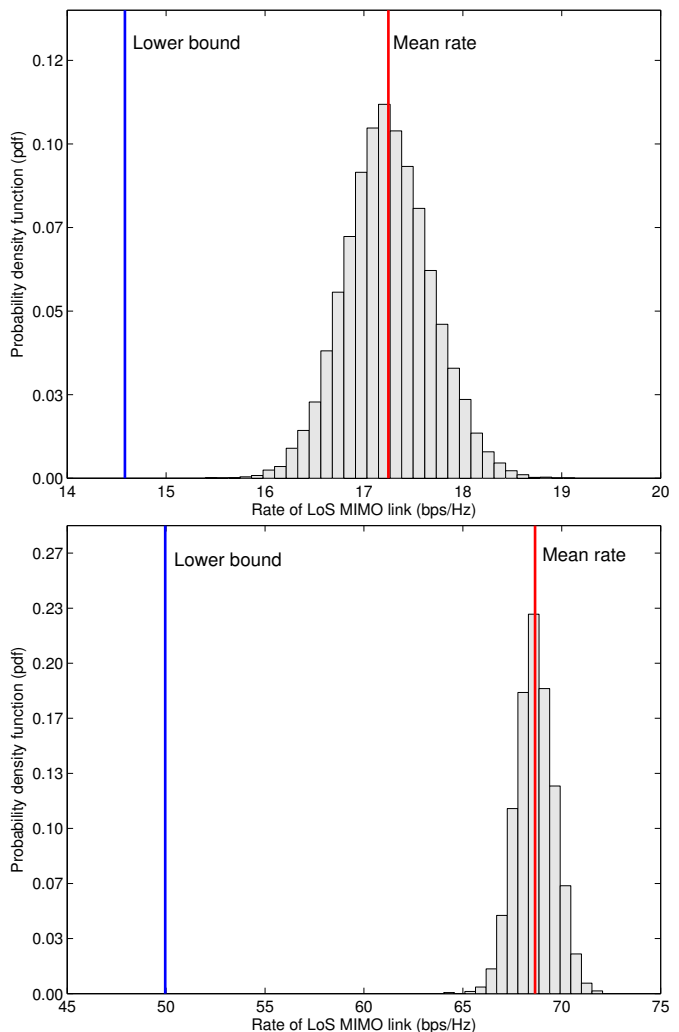


Fig. 6. Probability density function (pdf) of the LoS MIMO link rate. (first) 64 antennas, and (second) 256 antennas. Also plotted are the mean rate and the lower bound on the mean rate.

make each one of these mini-networks scalable. By bounding the interference power as in the proof of Theorem 3, it is easy to show that the mini-networks can operate simultaneously, thereby achieving scalability of the whole network. A tiny sample of related works in this direction is [22], [30].

VII. CONCLUDING REMARKS

Wireless backhaul for current cellular networks is quickly becoming a necessity, especially in the context of urban small cell deployments. Two likely features of future backhaul networks are: (i) higher transmission frequencies (mm-waves), and consequently, (ii) large number of transmit antennas per BS. We modeled this network as a multi-antenna random extended network, where the number of antennas per BS can scale as some arbitrary function of the total number of BSs. Using geometric arguments, we derived an information theoretic upper bound on the capacity of this network, and consequently, a lower bound on the number of antennas per BS required for throughput scalability. We further compared the scalability requirements of two competing strategies of

interest. While the first minimizes the number of hops by forming thin beams and ideally achieving full beamforming gain over each hop, the other communicates data from source to its destination through many short hops, each achieving full multiplexing gain. Although it may seem intuitive at first to minimize the number of hops by forming narrow beams and hence minimizing interference to neighboring nodes, we show that the short hop strategy is significantly better. The achievability result for the short hop strategy is derived under LoS propagation while carefully accounting for the fundamental limits on the spatial DoFs of a LoS MIMO channel.

From the practical system design viewpoint, some important guidelines can be drawn from our analysis. In the large-scale MIMO regime advocated by the short-hop architecture, large spectral efficiency per link can be achieved with very simple coding and modulation schemes. For example, in the numerical calculations of Section VI we achieve 17 bps/Hz with 64 antennas per BS, which implies 0.266 bit/stream. This is easily achievable by using QPSK concatenated with power capacity-approaching binary (Turbo/LDPC) codes. Reuse results in Fig. 5 highlight the importance of well-planned reuse, which along with the *optimal* distributed power control forms an important area of future work under the umbrella of self-organizing backhaul networks. Finally, the simple extension of our analysis to the case where some BSs have access to wired backhaul (Section VI) yields an appealing system design tradeoff: the per-BS hardware complexity (number of antennas) can be reduced from $n^{1/2}$ to $n^{(1-\beta)/2}$ in the presence of n^β wired backhaul BSs uniformly placed in the network.

Concrete directions for future work may include: (i) develop a communication theoretic channel model that captures the DoFs bottleneck of electromagnetic propagation [13] without necessarily assuming LoS; (ii) rigorize the approximation argument in [26], [29] which is instrumental for the sharp characterization of the DoFs of LoS MIMO channels; and (iii) performance evaluation of backhaul wireless networks under realistic population conditions in order to assess if the guidelines obtained from scaling laws reflect in actual system gains.

APPENDIX

A. Proof of Lemma 1

Recall that for $X \sim \text{Pois}(\lambda_b)$, we have the following Chernoff bound for all $x > \lambda_b$

$$\mathbb{P}(X \geq x) \leq \exp(-\lambda_b) \left(\frac{\exp(1)\lambda_b}{x} \right)^x. \quad (34)$$

Please refer to [7, Appendix II] or [31, Theorem 5.4] for further details. Also note that $N(\mathcal{A}) \sim \text{Pois}(\lambda_b|\mathcal{A}|)$ [32]. Specializing the above bound for $N(\mathcal{A})$ and substituting $x = 2\lambda_b|\mathcal{A}|$, we get

$$\mathbb{P}(N(\mathcal{A}) \geq 2\lambda_b|\mathcal{A}|) \leq e^{-\lambda_b|\mathcal{A}|} \left(\frac{e}{2} \right)^{2\lambda_b|\mathcal{A}|} = \left(\frac{e}{4} \right)^{\lambda_b|\mathcal{A}|}, \quad (35)$$

from which (2) follows by the fact that $\frac{e}{4} < 1$. Similarly, for all $x < \lambda_b$, we have the same Chernoff bound

$$\mathbb{P}(X \leq x) \leq \exp(-\lambda_b) \left(\frac{\exp(1)\lambda_b}{x} \right)^x. \quad (36)$$

Again specializing this bound for $N(\mathcal{A})$ and substituting $x = \frac{\lambda_b}{2}|\mathcal{A}|$, we get

$$\mathbb{P}\left(N(\mathcal{A}) \leq \frac{\lambda_b}{2}|\mathcal{A}|\right) \leq e^{-\lambda_b|\mathcal{A}|} (2e)^{\frac{\lambda_b}{2}|\mathcal{A}|} = \left(\frac{2}{e} \right)^{\frac{\lambda_b}{2}|\mathcal{A}|}, \quad (37)$$

from which (3) follows by the fact that $\frac{2}{e} < 1$. ■

B. Proof of Lemma 2

For notational simplicity, define $\kappa = \frac{\alpha}{2} - 2 - \log_n \Psi(n)$, and assume it to be positive, which implies that this derivation is applicable for $\alpha > 2(2 + \log_n \Psi(n))$. This condition will be required for the Taylor expansion of the log terms. The summation term C_s can be expressed as

$$\begin{aligned} C_s &= \sum_{i=1}^{\log \frac{\sqrt{n}}{2}} \frac{n}{e^i} \log(1 + Pn^{2-\frac{\alpha}{2}} \Psi(n) 2^\alpha e^{i\alpha}) \\ &= \sum_{i=1}^{\log \frac{\sqrt{n}}{2}} \frac{n}{e^i} \log\left(1 + P2^\alpha \frac{e^{i\alpha}}{n^\kappa}\right) \\ &= \sum_{i=1}^{\frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2}} \frac{n}{e^i} \log\left(1 + P2^\alpha \frac{e^{i\alpha}}{n^\kappa}\right) + \\ &\quad \sum_{i=\frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2} + 1}^{\log \frac{\sqrt{n}}{2}} \frac{n}{e^i} \log\left(1 + P2^\alpha \frac{e^{i\alpha}}{n^\kappa}\right) = C_{s_1} + C_{s_2}. \end{aligned} \quad (38)$$

Since the constant $P2^\alpha$ in the log terms of both C_{s_1} and C_{s_2} is independent of n and hence does not impact scaling of these terms, we will ignore it in the following discussion for notational simplicity. Using Taylor series expansion for log term, C_{s_1} can now be expressed as

$$\begin{aligned} C_{s_1} &= \sum_{i=1}^{\frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2}} \frac{n}{e^i} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \frac{e^{ki\alpha}}{n^{k\kappa}} \\ &= n \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \frac{1}{n^{k\kappa}} \sum_{i=1}^{\frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2}} e^{i(k\alpha-1)}. \end{aligned} \quad (39)$$

The summation with respect to i can be computed as

$$\begin{aligned} \sum_{i=1}^{\frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2}} e^{i(k\alpha-1)} &= \frac{e^{k\alpha-1}}{e^{k\alpha-1} - 1} \left(\frac{n^{(k\alpha-1)\kappa/\alpha}}{2^{2(k\alpha-1)\kappa/\alpha}} - 1 \right) \\ &\stackrel{(a)}{\leq} m \left(\frac{n^{(k\alpha-1)\kappa/\alpha}}{2^{2(k\alpha-1)\kappa/\alpha}} - 1 \right), \end{aligned} \quad (40)$$

where (a) follows by the fact that $\frac{e^{k\alpha-1}}{e^{k\alpha-1}-1}$ is uniformly upper bounded by some positive constant m . Substituting (40) back in (39), we get

$$C_{s_1} \leq n \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \frac{1}{n^{k\kappa}} m \left(\frac{n^{(k\alpha-1)\kappa/\alpha}}{2^{2(k\alpha-1)\kappa/\alpha}} - 1 \right)$$

$$= C_{s_{11}} + C_{s_{12}}. \quad (41)$$

Ignoring again the constants, $C_{s_{11}}$ can be upper bounded as

$$C_{s_{11}} = n^{1-\frac{\kappa}{\alpha}} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = n^{1-\frac{\kappa}{\alpha}} \log 2 = O(n^{1-\frac{\kappa}{\alpha}}). \quad (42)$$

Similarly, $C_{s_{12}}$ can be upper bounded as

$$C_{s_{12}} = n \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \frac{1}{n^{k\kappa}} = n \log(1 + n^{-\kappa}) = O(n^{1-\kappa}). \quad (43)$$

Substituting (42) and (43) back in (41), we get

$$C_{s_1} \leq O(n^{1-\frac{\kappa}{\alpha}}) + O(n^{1-\kappa}) = O(n^{1-\frac{\kappa}{\alpha}}), \quad (44)$$

where the last equality follows from the fact that $\alpha > 2$. We now turn our attention to C_{s_2} in (38), where we again ignore the constants. It can be expressed as

$$\begin{aligned} C_{s_2} &= \sum_{i=\frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2} + 1}^{\log \frac{\sqrt{n}}{2}} \frac{n}{e^i} \log \left(1 + \frac{e^{i\alpha}}{n^\kappa} \right) \\ &= \sum \frac{n}{e^i} \log \left(e^{i\alpha} \cdot \frac{1}{n^\kappa} \cdot \left(1 + \frac{n^\kappa}{e^{i\alpha}} \right) \right) \\ &= \sum \frac{n\alpha i}{e^i} - \sum \frac{n\kappa}{e^i} \log n + \sum \frac{n}{e^i} \log \left(1 + \frac{n^\kappa}{e^{i\alpha}} \right) \\ &= C_{s_{21}} - C_{s_{22}} + C_{s_{23}}, \end{aligned} \quad (45)$$

where the limits of the summation for all the terms are the same as the first equation. We now look at the three terms separately starting with $C_{s_{21}}$

$$\begin{aligned} C_{s_{21}} &= n\alpha \sum_{i=\frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2} + 1}^{\log \frac{\sqrt{n}}{2}} \frac{i}{e^i} \\ &\stackrel{(a)}{\leq} O(\sqrt{n}) + O(n^{1-\frac{\kappa}{\alpha}}) \stackrel{(b)}{=} O(n^{1-\frac{\kappa}{\alpha}}), \end{aligned} \quad (46)$$

where (a) follows by computing the summation directly, and (b) from the fact that $\frac{\kappa}{\alpha} = \frac{1}{2} - \frac{1}{\alpha}(2 + \log_n \Psi(n)) \leq \frac{1}{2}$. Similarly, the second term can be expressed as

$$\begin{aligned} C_{s_{22}} &= n\kappa \log(n) \sum_{i=\frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2} + 1}^{\log \frac{\sqrt{n}}{2}} e^{-i} \stackrel{(a)}{\leq} O(\sqrt{n} \log(n)) + \\ &O(n^{1-\frac{\kappa}{\alpha}} \log(n)) = O(n^{1-\frac{\kappa}{\alpha}} \log(n)), \end{aligned} \quad (47)$$

where (a) again follows by computing the summation directly and the final result by the fact that $\kappa < \frac{\alpha}{2}$. Now we come to the final term $C_{s_{23}}$, which can be expressed in terms of the Taylor series, $\log(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k$, $|x| \leq 1$, as

$$\begin{aligned} C_{s_{23}} &= n \sum_{i=\frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2} + 1}^{\log \frac{\sqrt{n}}{2}} \frac{1}{e^i} \log \left(1 + \frac{n^\kappa}{e^{i\alpha}} \right) \\ &= n \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} n^{k\kappa} \sum_{i=\frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2} + 1}^{\log \frac{\sqrt{n}}{2}} \frac{1}{e^{i(k\alpha+1)}}. \end{aligned} \quad (48)$$

Note that the above expansion of log term only holds when $\frac{n^\kappa}{e^{i\alpha}} < 1$. Since $\frac{n^\kappa}{e^{i\alpha}}$ is a decreasing function of i , it is sufficient to show that the condition holds for $i = \frac{2\kappa}{\alpha} \log \frac{\sqrt{n}}{2} + 1$. For this value of i , it is easy to show that $\frac{n^\kappa}{e^{i\alpha}} = \frac{2^{2\kappa}}{e^\alpha} \stackrel{(a)}{<} \left(\frac{2}{e}\right)^\alpha \stackrel{(b)}{<} 1$, where (a) follows from the fact that $2\kappa < \alpha$, and (b) is due to $\frac{2}{e} < 1$ and $\alpha > 2$. Now, the summation with respect to i can be expressed as

$$\begin{aligned} &\frac{e^{-k\alpha-1}}{1-e^{-k\alpha-1}} \frac{1}{e^{\frac{k\alpha+1}{\alpha} 2\kappa \log \frac{\sqrt{n}}{2}}} \left[1 - \frac{e^{k\alpha+1}}{e^{(k\alpha+1)(1-\frac{2\kappa}{\alpha}) \log \frac{\sqrt{n}}{2}}} \right] \\ &= \frac{e^{-k\alpha-1}}{1-e^{-k\alpha-1}} \left(\frac{4}{n} \right)^{k\kappa + \frac{\kappa}{\alpha}} - \frac{1}{1-e^{-k\alpha-1}} \left(\frac{4}{n} \right)^{\frac{\alpha k+1}{2}}. \end{aligned} \quad (49)$$

Using the fact that both $\frac{e^{-k\alpha-1}}{1-e^{-k\alpha-1}}$ and $\frac{1}{1-e^{-k\alpha-1}}$ are upper bounded uniformly by positive constants, and ignoring the constants that do not impact scaling laws, $C_{s_{23}}$ can be upper bounded by the sum of the following two terms

$$C_{s_{231}} = n^{1-\frac{\kappa}{\alpha}} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = n^{1-\frac{\kappa}{\alpha}} \log 2 = O(n^{1-\frac{\kappa}{\alpha}}) \quad (50)$$

$$\begin{aligned} C_{s_{232}} &= \sqrt{n} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} n^{(\kappa-\frac{\alpha}{2})k} \stackrel{(a)}{=} \sqrt{n} \log(1 + n^{\kappa-\frac{\alpha}{2}}) \\ &= O\left(\frac{\sqrt{n}}{n^{\frac{\alpha}{2}-\kappa}}\right) = O(\sqrt{n}), \end{aligned} \quad (51)$$

where (a) follows from the fact that since $\kappa - \frac{\alpha}{2} = -2 - \log_n \Psi(n) < 0$, $n^{k(\kappa-\frac{\alpha}{2})} < 1$, followed by using the appropriate Taylor series expansion. Combining these two results, we get the following upper bound on $C_{s_{23}}$

$$C_{s_{23}} \leq O(n^{1-\frac{\kappa}{\alpha}}) + O(\sqrt{n}) = O(n^{1-\frac{\kappa}{\alpha}}). \quad (52)$$

Substituting (46), (47) and (52) in (45), C_{s_2} can be upper bounded as

$$\begin{aligned} C_{s_2} &= C_{s_{21}} - C_{s_{22}} + C_{s_{23}} \\ &\leq O(n^{1-\frac{\kappa}{\alpha}}) + O(n^{1-\frac{\kappa}{\alpha}} \log(n)) + O(n^{1-\frac{\kappa}{\alpha}}) \\ &= O(n^{1-\frac{\kappa}{\alpha}} \log n). \end{aligned} \quad (53)$$

Now combining (44) and (53), C_s can be upper bounded as

$$\begin{aligned} C_s &= C_{s_1} + C_{s_2} = O(n^{1-\frac{\kappa}{\alpha}}) + O(n^{1-\frac{\kappa}{\alpha}} \log(n)) \\ &= O(n^{1-\frac{\kappa}{\alpha}} \log n), \end{aligned} \quad (54)$$

which completes this proof. \blacksquare

C. Proof of Lemma 3

Let the source-destination separation of a randomly chosen pair be D_{B_n} . Denote the locations of the source and destination BSs for this pair by $(X_s, Y_s) \in \mathbb{R}^2$ and $(X_d, Y_d) \in \mathbb{R}^2$, where X_s and Y_s are i.i.d. random variables uniformly distributed in $[0, \sqrt{n}]$. Similarly X_d and Y_d are i.i.d. and uniformly distributed in $[0, \sqrt{n}]$. The cumulative distribution function (CDF) of D_{B_n} is

$$\begin{aligned} \mathbb{P}(D_{B_n} \leq z) &= \mathbb{P}((X_s - X_d)^2 + (Y_s - Y_d)^2 \leq z^2) \\ &\leq \mathbb{P}((X_s - X_d)^2 \leq z^2) \\ &= \mathbb{P}(-z + X_d \leq X_s \leq z + X_d) \end{aligned}$$

$$\leq \frac{2z}{\sqrt{n}}, \quad (55)$$

where the inequality in the last step is because we ignored the restrictions on the range of X_s . Clearly $\lim_{n \rightarrow \infty} \mathbb{P}(D_{B_n} \leq n^{\frac{1}{2}-\epsilon}) = 0$ for $\epsilon > 0$, from which the result follows. ■

D. Proof of Lemma 4

First note that the exponential term can be equivalently expressed as $\exp(-j2\pi D\Psi(n)) = \exp(-j2\pi X)$, where $X = D\Psi(n) \bmod 1$, which clearly lies in $[0, 1]$. It is enough to show that as $\Psi(n)$ grows large, X tends to a uniform distribution between $[0, 1]$. Therefore for $x \in [0, 1]$

$$\begin{aligned} \mathbb{P}(X \leq x) &= \sum_{i=0}^{\infty} \mathbb{P}(i \leq D\Psi(n) \leq i+x) \\ &= \sum_{i=0}^{\infty} \mathbb{P}\left(\frac{i}{\Psi(n)} \leq D \leq \frac{i+x}{\Psi(n)}\right) \end{aligned} \quad (56)$$

$$\stackrel{(a)}{=} x \sum_{i=0}^{\infty} \frac{1}{\Psi(n)} f_D\left(\frac{i}{\Psi(n)}\right) = x, \quad (57)$$

where (a) and its subsequent step hold under $\Psi(n) \rightarrow \infty$ and $f_D(\cdot)$ denotes the probability density function of D . Clearly X is uniformly distributed in $[0, 1]$. ■

E. Supplement to Section V-B: Achievability of $R_{\text{sh}}(n) = \Omega(\sqrt{n})$ using $\Psi(n) = \sqrt{n}$.

Recall that in Section V-B we showed that $\Psi(n) = \sqrt{n}$ is sufficient to achieve $\mathbb{E}[R_{\text{sh}}(n)] = \Omega(\sqrt{n})$. However, this does not imply scalability w.h.p. for a given realization of the BS antenna locations. To address this, we numerically evaluated the distribution of achievable rate $R_{\text{sh}}(n)$ in Section VI and showed that the achievable rate (i) concentrates around its mean, and (ii) is lower bounded w.h.p. by our average rate lower bound. In this Appendix, we present a proof idea based on the analysis and insights of [29] on LoS MIMO channels, where it was shown that the number of significant singular values of the LoS MIMO channel matrix is $\frac{a}{\lambda d}$, where d is the transmitter-receiver separation, and they are roughly of the same order.

Using the fact that the rate lower bound in (25) is decreasing with respect to \mathbf{R} on the cone of positive semidefinite Hermitian matrices, and that $\lambda_{\max}(\mathbf{R})\mathbf{I} \geq \mathbf{R}$, where $\lambda_{\max}(\cdot)$ is the spectral radius, we have the lower bound

$$R_{\text{sh}}(n) \geq \log \det \left(\mathbf{I} + \frac{P}{\Psi(n)} \frac{1}{\lambda_{\max}(\mathbf{R})} \mathbf{H}\mathbf{H}^\dagger \right). \quad (58)$$

From the definition of \mathbf{R} given by (26), we have

$$\lambda_{\max}(\mathbf{R}) \leq 1 + \frac{P}{\Psi(n)} \sum_{i \in \mathcal{I}} \lambda_{\max}(\mathbf{H}^{(i)}\mathbf{H}^{(i)\dagger}) \quad (59)$$

According to the analysis of [29], $\mathbf{H}^{(i)}\mathbf{H}^{(i)\dagger}$ has $\frac{a}{\lambda d^{(i)}}$ significant eigenvalues, each with value (in order) $\frac{\left[\min\left\{1, (d^{(i)})^{-\frac{\alpha}{2}}\right\}\right]^2 \Psi(n)^2 \lambda d^{(i)}}{a}$, where we used the fact that the sum of the eigenvalues is equal to $\text{Tr}(\mathbf{H}^{(i)}\mathbf{H}^{(i)\dagger}) = \left[\min\left\{1, (d^{(i)})^{-\frac{\alpha}{2}}\right\}\right]^2 \Psi(n)^2$. To ensure

full DoF for the desired link, we have $\frac{a}{\lambda d_{\max}} = \Psi(n)$, implying that the eigenvalues of $\mathbf{H}^{(i)}\mathbf{H}^{(i)\dagger}$ can be expressed (in order) as $\left[\min\left\{1, (d^{(i)})^{-\frac{\alpha}{2}}\right\}\right]^2 \Psi(n) d^{(i)} d_{\max}^{-1}$. Substituting this in (59) in place of $\lambda_{\max}(\mathbf{H}^{(i)}\mathbf{H}^{(i)\dagger})$, we get

$$\lambda_{\max}(\mathbf{R}) \leq 1 + P \sum_{i \in \mathcal{I}} \left[\min\left\{1, (d^{(i)})^{-\frac{\alpha}{2}}\right\}\right]^2 d^{(i)} d_{\max}^{-1}, \quad (60)$$

which is the same as the constant term of $\mathbb{E}[\mathbf{R}]$ in (30), except that the above expression contains an extra distance term $d^{(i)}$. Following the same procedure as for (30), we can upper bound $\lambda_{\max}(\mathbf{R})$ by a constant, but due to the presence of extra $d^{(i)}$, this will be valid for $\alpha > 3$. Note that the above arguments are based on the conjecture that all the eigenvalues of $\mathbf{H}^{(i)}\mathbf{H}^{(i)\dagger}$ are “exactly” the same (in order) (hence equal to $\lambda_{\max}(\mathbf{H}^{(i)}\mathbf{H}^{(i)\dagger})$), which is motivated by the discussion in [29]. To rigorize this proof idea, we need to prove this conjecture and show that the element wise approximations made in [29] imply convergence to the same limit of the asymptotic distribution of the actual LoS MIMO channel matrix. This is a promising area for future work.

REFERENCES

- [1] H. S. Dhillon and G. Caire, “Information theoretic upper bound on the capacity of wireless backhaul networks,” in *Proc., IEEE Intl. Symposium on Information Theory*, Honolulu, HI, Jul. 2014.
- [2] —, “Scalability of line-of-sight massive MIMO mesh networks for wireless backhaul,” in *Proc., IEEE Intl. Symposium on Information Theory*, Honolulu, HI, Jul. 2014.
- [3] T. Q. S. Quek, G. de la Roche, I. Güvenç, and M. Kountouris, Eds., *Small Cell Networks: Deployment, PHY Techniques, and Resource Management*. Cambridge University Press, Jun. 2013.
- [4] Small Cell Forum, “Backhaul technologies for small cells”, white paper, document 049.02.01, Dec. 2013.
- [5] Cambridge Communication Systems, “A backhaul solution for organic small-cell growth,” white paper, 2014, available online: ccsl.com/resources/.
- [6] P. Gupta and P. R. Kumar, “The capacity of wireless networks,” *IEEE Trans. on Info. Theory*, vol. 46, no. 2, pp. 388 – 404, Mar. 2000.
- [7] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, “Closing the gap in the capacity of wireless networks via percolation theory,” *IEEE Trans. on Info. Theory*, vol. 53, no. 3, pp. 1009 – 1018, Mar. 2007.
- [8] G. R. MacCartney Jr., J. Zhang, S. Nie, and T. S. Rappaport, “Path loss models for 5G millimeter wave propagation channels in urban microcells,” in *Proc., IEEE Globecom*, Atlanta, GA, Dec. 2013.
- [9] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!” *IEEE Access*, vol. 1, pp. 335 – 349, 2013.
- [10] A. Adhikary, E. A. Safadi, M. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, “Joint spatial division and multiplexing for mm-wave channels,” 2013, *arXiv preprint*, available online: arxiv.org/abs/1312.2045.
- [11] L.-L. Xie and P. R. Kumar, “A network information theory for wireless communication: Scaling laws and optimal operation,” *IEEE Trans. on Info. Theory*, vol. 50, no. 5, pp. 748 – 767, May 2004.
- [12] A. Özgür, O. Lévêque, and D. N. C. Tse, “Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks,” *IEEE Trans. on Info. Theory*, vol. 53, no. 10, pp. 3549 – 3572, Oct. 2007.
- [13] M. Franceschetti, D. Migliore, and P. Minero, “The capacity of wireless networks: Information-theoretic and physical limits,” *IEEE Trans. on Info. Theory*, vol. 55, no. 8, pp. 3413 – 3424, Aug. 2009.
- [14] S.-H. Lee and S.-Y. Chung, “Capacity scaling of wireless ad hoc networks: Shannon meets Maxwell,” *IEEE Trans. on Info. Theory*, vol. 58, no. 3, pp. 1702 – 1715, Mar. 2012.

- [15] A. Özgür, O. Lévêque, and D. N. C. Tse, "Spatial degrees of freedom of large distributed MIMO systems and wireless ad hoc networks," *IEEE Journal on Sel. Areas in Commun.*, vol. 31, no. 2, pp. 202 – 214, Feb. 2013.
- [16] A. Özgür, R. Johari, D. N. C. Tse, and O. Lévêque, "Information-theoretic operating regimes of large wireless networks," *IEEE Trans. on Info. Theory*, vol. 56, no. 1, pp. 427 – 437, Jan. 2010.
- [17] V. Gambiroza, B. Sadeghi, and E. Knightly, "End-to-end performance and fairness in multihop wireless backhaul networks," in *Proc., ACM MobiCom*, Philadelphia, PA, Sep. 2004.
- [18] M. Cao, X. Wang, S.-J. Kim, and M. Madihan, "Multi-hop wireless backhaul networks: A cross-layer design paradigm," *IEEE Journal on Sel. Areas in Commun.*, vol. 25, no. 4, pp. 738 – 748, May 2007.
- [19] M. Anastasopoulos, P.-D. Arapoglou, R. Kannan, and P. Cottis, "Adaptive routing strategies in IEEE 802.16 multi-hop wireless backhaul networks based on evolutionary game theory," *IEEE Journal on Sel. Areas in Commun.*, vol. 26, no. 7, pp. 1218 – 1225, Sep. 2008.
- [20] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. on Commun.*, vol. 61, no. 10, pp. 4391 – 4403, Oct. 2013.
- [21] M. Franceschetti, "A note on Lévêque and Telatar's upper bound on the capacity of wireless ad hoc networks," *IEEE Trans. on Info. Theory*, vol. 53, no. 9, pp. 3207 – 3211, Sep. 2007.
- [22] W.-Y. Shin, S.-W. Jeon, N. Devroye, M. H. Vu, S.-Y. Chung, Y. H. Lee, and V. Tarokh, "Improved capacity scaling in wireless networks with infrastructure," *IEEE Trans. on Info. Theory*, vol. 57, no. 8, pp. 5088 – 5102, Aug. 2011.
- [23] M. Franceschetti and R. Meester, *Random Networks for Communication: From Statistical Physics to Information Systems*. Cambridge University Press, 2007.
- [24] O. Lévêque and I. E. Telatar, "Information-theoretic upper bounds on the capacity of large extended ad hoc wireless networks," *IEEE Trans. on Info. Theory*, vol. 51, no. 3, pp. 858 – 865, Mar. 2005.
- [25] F. Bøhagen, P. Orten, and G. E. Øien, "Design of optimal high-rank line-of-sight MIMO channels," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 4, pp. 1420 – 1425, Apr. 2007.
- [26] M. Desgroseilliers, O. Lévêque, and E. Preissmann, "Spatial degrees of freedom of MIMO systems in line-of-sight environment," in *Proc., IEEE Intl. Symposium on Information Theory*, Istanbul, Turkey, Jul. 2013.
- [27] A. Bleicher, "The 5G phone future: Samsung's millimeter-wave transceiver technology could enable ultrafast mobile broadband by 2020," *IEEE Spectrum*, vol. 50, no. 7, pp. 15 – 16, Jul. 2013.
- [28] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York: McGraw-Hill, 2008.
- [29] M. Desgroseilliers, O. Lévêque, and E. Preissmann, "Partially random matrices in line-of-sight wireless networks," in *Proc., IEEE Asilomar*, Pacific Grove, CA, Nov. 2013.
- [30] A. Zemplianov and G. de Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE Journal on Sel. Areas in Commun.*, vol. 23, no. 3, pp. 657 – 667, Mar. 2005.
- [31] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.
- [32] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks, Volume I – Theory*. NOW: Foundations and Trends in Networking, 2009.